How Long Do Treatment Effects Last? Persistence and Durability of a Descriptive Norms Intervention's Effect on Energy Conservation

How Long Do Treatment Effects Last? Persistence and Durability of a Descriptive Norms Intervention's Effect on Energy Conservation

Faculty Research Working Paper Series

Hunt Allcott
New York University

Todd Rogers
Harvard Kennedy School

October 2012
RWP12-045

How Long Do Treatment Effects Last?

Persistence and Durability of a Descriptive Norms Intervention's Effect on Energy Conservation

Hunt Allcott

Economics Department, New York University

Todd Rogers

Center for Public Leadership, Harvard Kennedy School

(Currently under review)

We thank Sendhil Mullainathan, Eldar Shafir, Francesca Gino, and Mike Norton for helpful conversations. Thanks to Tyler Curtis, Lisa Danz, Rachel Gold, Arkadi Gerney, Marc Laitin, Laura Lewellyn, and many others at OPOWER for sharing data and insight with us. Thanks to Carly Robinson for research help. We are grateful to the Sloan Foundation for financial support of our research on the economics of energy efficiency. Stata code for replicating the analysis is available from Hunt Allcott's website.

ABSTRACT(149 words):

Behavioral decision research has profoundly changed our understanding of decision-making. Recent research has begun to explore how behavioral insights can influence behavior in the world, at scale. This work often involves field experiments studying outcomes over short time windows. We study a descriprive social norms intervention's impact on household energy usage continuously over 39 to 49 months. Our two field experiments (N=155,000 households) each have three conditions: untreated control, continued treatment, and treatment that is subsequently discontinued. We find that continued treatment reduces energy usage over the entire period ("durability"). Further, after treatment is discontinued, a sizable energy use reduction persists ("persistence"). Finally, continued treatment generates a greater impact over time than discontinued treatment, showing that continued treatment exerts incremental influence on behavior over and above persistence. We discuss implications, describe how long-term persistence can occur, and argue that future behavioral decision research should address long-term effects of interventions.

Since the work of Herb Simon in the late 1950's, behavioral decision researchers have developed a sophisticated understanding of human decision-making. This work has shown how and when people are not perfectly rational, and the systematic patterns in their judgments and decisions (e.g., Kahneman and Tversky 1979; Gilovich, Griffin, and Kahneman 2002). For sensible and practical reasons, this research has tended to study decisions in laboratories, surveys, hypothetical scenarios, and artificial field settings (for review see Baumeister, Vohs and Funder 2007). In recent years, however, there has been a move toward extending these behavioral insights by using large-scale natural field experiments (e.g., Schultz et al. 2007; Madrian and Shea 2000). This approach – extending behavioral decision research using field experiments – has had a decidedly prescriptive thrust (Thaler and Sunstein 2003, 2009; Camerer, Loewenstein and Prelec 2003); in addition to deepening our understanding of human behavior, it has tended to examine contexts and interventions that help us understand and influence pressing societal problems. The preponderance of this field research has studied brief treatments, and measured outcomes that occur immediately after, or concurrent with, the treatment. Scant work has examined the dynamics of treatment effects as treatments are sustained over time (what we term "durability"), and whether treatment effects survive after treatments are discontinued (what we term "persistence"). The present manuscript explores the durability and persistence over three to four years of an intervention aimed at reducing people's energy usage by leveraging people's conformity to descriptive social norms.

While there have been many field experiments looking at how behavioral theories affect real behavior in the field (e.g., Gneezy and List 2006; Bertrand and Mullainathan 2003; Ashraf, Karlan, and Yin 2006; Paluck 2009; Nickerson and Rogers 2010; Fryer, Levitt, List, and Sadoff 2012), a large fraction examine outcomes measured only once and usually very shortly after treatment is administered, and therefore do not examine the dynamics and survival of treatment effects over the long-term. There are many potential explanations for this, including the possibility that long-term effects have not been critical to the core research questions being investigated, practical considerations concerning the timeline and incentives for the publication of academic research, or motivated non-reporting of null long-term effects, to name a few.

Occasionally, studies do report having examined long-term persistence or durability, and they often show rapid decay of treatment effects. In a typical example in the weight-loss domain, John et al. showed that including a commitment device involving risking one's own money in a weight-loss program resulted in significantly more weight loss during an eight month program, but that the weight was then regained over the next four weeks (2011). Similarly, in the smoking cessation domain, a recent meta-analysis of seventeen rigorous studies of incentives and competitions to induce long-term smoking cessation found no average long-term effect (Cahill and Perera 2008). That said, a handful of studies have examined and observed both short-term and somewhat longer-term effects (e.g., Charness and Gneezy 2009; Walton and Cohen 2011; Volpp et al. 2009; Feraro, Miranda, and Price 2011) and some researchers have begun wrestling with why long-term effects might occur (Yeager and Walton 2011).

The present experiments examine the long-term durability and persistence of a behavioral intervention that has been shown in multiple experiments to reduce energy usage (descriptive social norms). The design of the experiments allows us to study the long-term persistence and durability of the energy-reducing treatment. These findings show that behavioral interventions can yield long-term behavior change that is additive when treatments are continued, and that persist after they are discontinued.

SOCIAL NORMS AND BEHAVIOR

Social norms are often characterized as being of two types, injunctive and descriptive. Injunctive norms describe peoples' beliefs about what others think they should do (e.g., ''You should not waste energy"), while descriptive norms describe peoples' beliefs about what others actually do (e.g., ''Most people use a lot of energy"). Both types of norms, when made salient, tend to encourage norm-consistent behavior (see Reno, Cialdini, and Kallgren 1993). This implies that including descriptive social norms in persuasive appeals can motivate behavior – assuming that the norm is in the preferred direction (Cialdini et al. 2006). Descriptive social norms have been shown to affect stealing of petrified wood from the forest floor (Cialdini et al. 2006), littering (Cialdini, Reno, and Kallgren 1990), towel reuse in hotels

(Goldstein, Cialdini, and Griskevicius 2008), retirement savings (Beshears et al. 2012), charitable giving (Frey and Meier 2004), and motivation to vote in elections (Gerber and Rogers 2009).

Two research projects investigating descriptive social norms are of special relevance to the present research. In one, households received two written messages left on their doors conveying how much energy they consumed relative to their neighbors (N = 290). The first message reported energy usage based on the previous week, and the second message reflected energy usage over the previous two weeks (Schultz et al. 2007). Since there is natural distribution across households of the amount of energy consumed by households, some households were truthfully told that they consumed more energy than their neighbors, while others were truthfully told that they consumed less energy than their neighbors. Theory involving descriptive social norms suggests that those who consumed more energy than their neighbors would decrease their energy usage as a result of receiving the treatment, whereas those who consumed less than their neighbors would consume more energy as a result of the treatment. The theory was supported, as the experimenters found that energy usage changed in the predicted directions two weeks after receiving the first message and three weeks after receiving the second and final message among households that received this descriptive information treatment. Half of the households received additional information accompanying their energy usage information: they received smiley faces if they consumed less than their neighbors, and frowny faces if they consumed more than their neighbors. These smiley and frowny faces reinforced the injunctive norm that consuming less energy is good. For households that used less than their neighbors, receiving the injunctive norm information eliminated the increase in energy use caused by the descriptive norm.

A second related project by the same research team, delivered four successive door-hangers to target households (N = 391) that conveyed either motivational messaging about why the households should perform energy saving behaviors (e.g., save money, good for environment, etc.), or messaging about how a large percentage of their neighbors perform specific energy saving behaviors (e.g., "99% of people in your community reported turning off unnecessary lights to save energy"). These four door-hangers were delivered over the course of one month, and energy meters were read during several of these door-hanger deliveries, including the first and last delivery. Energy meter readings showed that the descriptive social norm information significantly reduced energy usage over the course of the month of treatment compared those who received the motivational messaging; those who were assigned to receive the four descriptive norms door-hangers used 8.5 percent less energy than those who were assigned to the other conditions (Nolan et al. 2008). Energy meters were also read one month after treatment had been discontinued. The authors report that those assigned to the descriptive social norms treatment persisted in using less energy at the time of this final meter reading than those assigned to the other conditions (7.0 percent less energy used), though the difference was not statistically significant.

There are two features of this study that are worth noting in relation to the research to be reported in this manuscript. First, the measurement of long-term persistence is one month after the treatment is discontinued. In the experiments reported below we look at a much longer post-treatment time period to study treatment effect persistence (13-15 months and 19-22 months after treatment is discontinued). Second, given the rapid treatment effect decay observed in this study, 19% in one month, one might predict that a large-scale intervention to reduce energy use leveraging descriptive social norms would not be particularly persistent. In the experiments reported below we observe much less dramatic decay which enables substantial treatment effect persistence over a longer time period.

Given the effectiveness of descriptive social norms messaging for reducing energy use, and the relative ineffectiveness of other types of messaging, the private sector has commercialized this energy conservation strategy. Each year, utility companies spend billions of dollars on energy conservation programs (Allcott and Greenstone 2012), which has given rise to a growing sector focused on developing and selling interventions informed by behavioral science. Descriptive social norms are among the most effective – and cost effective – of these interventions (Allcott and Mullainathan 2011).

CONTEXT

More than half of US states have Energy Efficiency Portfolio Standards, which require utilities selling electricity or natural gas to also induce their consumers to reduce energy consumption by a small percentage each year. The company that deployed the treatment studied in these experiments, Opower, is a third party company that works with utilities to help satisfy these and related energy conservation goals. As of summer 2012, Opower's programs were being implemented at 70 utilities across the United States, and there were 8.4 million households in treatment and control groups. This makes Opower one of the largest sources of randomized field experiments ever studied. Allcott (2011) and Allcott and Mullainathan (2012) study several of Opower's sites, showing that the programs reduce energy use by 1.4 to 2.8 percent relative to control.

The "treatment" in the experiments reported in this manuscript entails mailing Opower's Home Energy Reports to consumers on a continuing basis, every month or every several months. The central feature of the Home Energy Report is descriptive social norm information: the household's energy use for a given time period is compared against a group of 100 nearby households that are of similar sizes and use the same fuel (natural gas or electricity) for heating. As demonstrated in Figure 1 (front), the descriptive social norm information compares the household's energy use to the mean neighbor as well as the 20th percentile of the distribution. In addition to these descriptive norms, the reports also include personalized feedback on energy usage and injunctive norm information: households that use less than the 20th percentile of their neighbor comparison group receive two "smiley face" emoticons, and households that use less than the mean receive one. This combination of descriptive and injunctive norms was directly motivated by the two studies described above by Nolan *et al.* (2008) and Schultz *et al.* (2007). The back page of the Home Energy Reports contains additional information, such as the energy conservation tips demonstrated in Figure 1 (back).

EXPERIMENTS

Basic Design. We analyze two experiments which have identical basic designs and which occurred at different sites. The basic design involves three conditions. Households assigned to the *discontinued* condition receive Home Energy Reports (monthly or quarterly, as described below) for around two years, then the treatments are discontinued and the energy usage of these households is observed for more than one additional year. Households assigned to the *continued* treatment condition receive Home Energy Reports (either monthly or quarterly, as described below) during the entirety of the experiment, which was ongoing at the time when the data used in this manuscript was compiled, 39 to 49 months after treatment began. Those households assigned to the *untreated control* condition do not receive any Home Energy Reports. Household energy usage for all three conditions is observed from two to three years before treatment began until May 1, 2012.

In both experiments, households assigned to the continued and discontinued conditions received either monthly or quarterly reports. In Experiment 1, households were randomly assigned to one of the two levels of treatment frequency, while in Experiment 2, households were assigned to monthly treatment if and only if their pre-treatment energy usage was above a threshold. In both experiments, households were randomly assigned to the continued, discontinued, and control conditions across both levels of treatment frequency. This means that the same proportion of households in each condition received treatment monthly and quarterly, and that these households are balanced on observable characteristics. While those who receive monthly treatment save more energy, on average, than those who receive treatment quarterly, the basic patterns of persistence and durability do not differ by treatment frequency. In our analysis, we therefore combine effects for both levels of frequency.

Experiment 1 Details. Experiment 1 occurs at a medium-sized investor-owned utility in a part of the Midwest with cold winters and mild summers. In Experiment 1, the entire population of residential consumers was potentially eligible for the experiment. To be included in the actual experiment universe, a customer needed to have a single-family home, at least 12 months of energy bills at their existing location, as well as a sufficient number of neighbors to construct the neighbor comparisons. There were several other technical restrictions that affected a small number of households: customers had to have valid names and addresses, no negative electricity meter reads, at least one meter read in the last three months, no significant gaps in usage history, and exactly one account per customer per location, and they could not be on special medical rate plans. Furthermore, a handful of utility staff were automatically enrolled in the reports and thus were excluded from the experiment universe. This experiment universe was randomized into the three conditions: untreated control, discontinued, and continued. Table 1 shows details about this condition assignment.

As shown in Table 2, we divide the data from the experiment into seven periods for our empirical analysis. In Experiment 1, the treatments began on February 1, 2009. The 12-month baseline period was defined to begin in the earliest month when essentially all households in the experiment universe had valid meter reads. In Experiment 1 there are four months between the end of the baseline period and the beginning of treatment. This forms the pre-treatment period in our analyses of this experiment, which always control for baseline period energy use. The "joint treatment period" is the period in the experiment when both the continued and discontinued conditions received reports. As in other experiments examining the impact of the Home Energy Reports, there is a rapid initial energy use reduction over the first few months among the households receiving treatment. We thus separate the "Joint Treatment Period" into an initial phase and a later phase.

Those households assigned to the discontinued condition stopped receiving treatment after February 1, 2011. We monitor the effects over the next 12 months after treatment is discontinued, and then present one measure of long-run persistence based on treatment effects 13 to 15 months after February 1, 2011.

Experiment 2 Details. Experiment 2 occurs at a large municipal utility in the Southwest with temperate winters and hot summers. The requirements to be in the experiment universe in Experiment 2 were similar to the requirements for Experiment 1: customers needed to have at least 12 months of valid historical energy bills as well as satisfy several other technical requirements. The utility's customer base was much larger, so Opower restricted the potentially-eligible universe to the set of Census tracts within the city to maximize the number of homes that would be actually eligible. Unlike in Experiment 1, the actual experiment universe in Experiment 2 was randomized into the three conditions at the "block batch group" level instead of the household level, where a block batch group is a set of two to three contiguous census blocks with approximately 50 to 100 homes. All analyses of this experiment cluster standard errors by block batch group to reflect this level of randomization. Table 1 provides details of the experiment universe for this experiment.

As shown in Table 2, we divide the data from the experiment into seven periods for our empirical analysis. In Experiment 2 the treatments began on April 1, 2008. In this experiment, the baseline period can begin April 1, 2006, and the pre-treatment period begins April 1, 2007. Households assigned to the discontinued condition stopped receiving treatment after July 1, 2010. To match Experiment 1, we present a measure of long-run persistence among households assigned to the discontinued condition as measured 13-15 months after treatment is discontinued. In Experiment 2, our sample also includes an additional six months beyond the 13-15 months of observation reported in Experiment 1.

Data collection. As part of their normal billing process, utility personnel at the sites where Experiments 1 and 2 took place visit households approximately once every month to read their electricity meters, which record cumulative electricity usage over time. The difference in cumulative usage between each meter read date is our primary dependent variable. We observe 2.8 million meter reads across the 72,000 households in Experiment 1, and 4.5 million meter reads across the 83,000 households in Experiment 2. Average baseline-period electricity use per household is around 30 kilowatt-hours per day in both experiments. For context, a typical incandescent lightbulb uses 0.3 kWh over five hours of usage, and a typical refrigerator might use 1.5 kilowatt-hours per day. As Table 1 shows, treatment and control, as well as continued and discontinued, are balanced on baseline usage in both experiments.

Attrition. There are two types of attrition in these experiments. First, 1.9 percent of households in experiment 1 and 2.6 percent of households in experiment two actively opted out of receiving treatments. We continue to observe energy usage for these households, and to exclude them from the regressions would generate imbalance between treatment and control. Following Allcott (2011), we continue to define a household that opts out as a treated household, meaning that the "treatment" in these experiments is defined as "being mailed a report or opting out." If one wished to define "treatment" as "being mailed a report" then our estimates would be intent-to-treat estimates. The second form of attrition is that households become "inactive" by moving or falling below minimum technical thresholds for electricity use or number of neighbors that can be used for constructing neighbor comparisons. This is more common: over the approximately four years during which these experiments occurred, 15.2 percent of households became inactive in Experiment 1, and 22.3 of households became inactive in Experiment 2, largely because they moved addresses. We do not observe energy usage for most customers after they become inactive. Therefore, even if we do observe a household's electricity bill after it becomes inactive, we drop data from inactive accounts once they become inactive.

Counter to our expectations, the inactive rates differ among households assigned to the two treatments and those assigned to the control in both experiments. In Experiment 1, those assigned to the treatment conditions are 0.51% less likely to become inactive (p=0.057), while in Experiment 2, those assigned to the treatment conditions are 1.1% more likely to become inactive (p=0.091). For a number of reasons, we are not very concerned with this. First, there is no theoretical reason to expect that the treatment makes households more or less likely to move, which suggests that the imbalance is a statistical fluke. Second, the p-values indicate that the differences are not highly statistically significant. Third, Allcott (2011) shows that this form of imbalance is uncommon in Opower experiments, and there is in fact no imbalance in earlier versions of the data from these same experiments. Fourth, the differences are small relative to the overall inactive rates, meaning that they should be unlikely to generate significant bias. Fifth, the sign of the imbalance is positive in one experiment and negative in the other, while our basic econometric results and qualitative conclusions are the same, meaning that the impact of attrition would somehow have to be exactly opposite in the two experiments in order to drive our qualitative conclusions. Sixth, and perhaps most convincingly for us, we re-ran all of our regressions after dropping any household that becomes inactive at any point. Not one of the coefficients changed in a statistically significant or economically meaningful way.

EMPIRICAL STRATEGY

We ask three basic research questions. First, do those in the continued condition show treatment effect durability over the life of the experiment? More precisely, do households in the continued condition use less energy than those in the untreated control condition through the life of the experiment?

Second, do those in the discontinued condition show treatment effect persistence after the treatment has been discontinued? More precisely, do households in the discontinued condition use less energy than those in the untreated control condition after the treatment has been discontinued?

Finally, the third question is conditional on finding treatment effect durability among households in the continued condition (first research question), and treatment effect persistence among households in the discontinued condition after treatment has been discontinued (second research question). If these do occur, does continued treatment increase the treatment effect above and beyond the persistence of treatment effect after treatment is discontinued? More precisely, after the joint treatment period, how much less energy do households in the continued treatment condition consume relative to households in the discontinued treatment condition?

To address the first and second research questions, define Y_{it} as electricity use by household i for meter read date t. Define P_t^p as an indicator variable for whether meter read date t falls within period p, where p indexes the periods listed in Table 2. Define T_i, D_i, and E_i as indicator variables for whether household i is in the treatment, discontinued, and continued groups, respectively. Define a set of month-by-year indicator variables μ_{mt}, where m indexes the months and years of the sample. Finally, define B_{imt} as household i's average daily electricity use for the meter read in the same calendar month as t during the baseline period. The first regression is:

$$
\begin{aligned}
Y_{it} &= \tau^0 T_i P_t^0 + \tau^1 T_i P_t^1 + \tau^2 T_i P_t^2 \\
&+ \alpha^3 D_i P_t^3 + \alpha^4 D_i P_t^4 + \alpha^5 D_i P_t^5 \\
&+ \gamma^3 E_i P_t^3 + \gamma^4 E_i P_t^4 + \gamma^5 E P_t^5 \\
&+ \sum_m (\mu_{mt} + \theta_{mt} B_{imt}) + \varepsilon_{it}
\end{aligned}
$$

The coefficients τ^0, τ^1, and τ^2 in this regression are the treatment effects during the pre-treatment and joint treatment periods. τ^0 should be zero, because treatment has not started, and τ^1 and τ^2 should be negative, reflecting a decrease in electricity use. The α and γ coefficients, respectively, reflect the treatment effects for the discontinued and continued groups relative to control. These measure persistence and durability, respectively.

To address the third question, we use a second regression:

$$
Y_{it} = \sum_{p=0}^{5} (\tau^p T_i + \delta^p D_i) P_t^p + \sum_m (\mu_{mt} + \theta_{mt} B_{imt}) + \varepsilon_{it}
$$

The δ^p coefficients measure the difference in electricity use between the continued and discontinued groups in period p. In the first three periods – pre-treatment, early joint treatment, and late joint treatment period – δ^p should be zero, as both the continuing and discontinued groups have received the same treatment. After that, we expect that δ may be weakly positive, reflecting higher electricity use in the discontinued group relative to the continued group after treatment is discontinued.

In all regressions, we cluster standard errors by household to address serial auto correlation, per Bertrand, Duflo, and Mullainathan (2004). We also weight the observations by the number of days in the billing period, although this makes effectively no difference because nearly all billing periods are very close to one month long.

RESULTS

Figures 2 and 3 plot the treatment effects over time for the continued and discontinued treatment groups in Experiments 1 and 2, respectively. The effects are estimated as three-month moving averages, controlling for baseline average usage within household. Both experiments show the same basic trends. In period 0, there is no effect, as the treatment has not yet begun. The effects increase in absolute value quickly for the first year before leveling out somewhat. Treatment effects are negative, as the program causes a decrease in energy use. Seasonality is important: the effects are larger in absolute value in the summer and winter compared to the “shoulder periods” in the spring and fall. After those in the discontinued condition stop receiving reports, their treatment effects weaken.

The figures illustrate that the intervention has durable effects over the 39 and 49 month periods that we observe: as long as treatment continues, the treatment effects are statistically significant. This affirmatively addresses our first research question. In fact, the effects appear to continually increase slightly. The effects are also persistent: the effects continue to be statistically significant among those assigned to the discontinued condition after the end of their treatment. This affirmatively addresses our second research question.

Table 3 presents our statistical tests of persistence and durability. As the graphs suggest, the treatment effects are statistically zero in the pre-treatment period and statistically negative over all post-treatment periods for both conditions in both experiments. The effects during the joint treatment period are very similar in the two experiments: -0.88 and -0.84 kilowatt-hours per day, respectively. These magnitudes are economically significant: they are equivalent to turning off about 15 standard 60-watt lightbulbs for one hour each day, and they represent 2.9 and 2.6 percent of baseline energy use in Experiments 1 and 2, respectively. The effects on those in the discontinued condition are also very similar across experiments in the first year after treatment is discontinued: -0.73 and -0.72 kwh/day.

Interestingly, however, the longer-run persistence differs across utilities. During the quarter beginning one year after the reports are discontinued, those in the discontinued condition in Experiment 1 conserve 0.40 kWh/day, compared to 0.67 kWh/day among those in Experiment 2. Table 4 presents our tests of differences in treatment effects between those in the continued and discontinued conditions. During the pre-treatment and joint treatment periods, the coefficients on D, which are the δ coefficients in Equation (2), are not statistically different than zero. This reflects the fact that those in the discontinued and continued conditions have the same treatment effects while they are receiving the same treatment. After treatment is discontinued for those in the discontinued condition, their electricity use rises relative to those in the continued condition. These δ coefficients over these later periods reflect the incremental effects of continuing the intervention. These coefficients affirmatively address our third research question: continued treatment increases the treatment effect above and beyond the persistence of treatment effect after treatment is discontinued.

DISCUSSION

Over the past half century behavioral decision research has made vast strides in understanding the underlying cognitive processes behind human decision making. In recent years this research has begun to examine how robust and potent this understanding can be in influencing actual behavior in the world.

This recent wave of research has often taken the form of field experiments targeting specific behaviors over relatively short windows of time. If behavioral decision research is to inform and strengthen interventions in the world, studies are needed of behavioral treatments that influence consequential behaviors over multiple years. In this manuscript we contribute to this work by examining how an intervention that is informed by behavioral decision research affects energy usage over many years. We report two field experiments examining an intervention to reduce energy usage involving 155,000 households. Both experiments illuminate three research questions. First, we find that continued administration of treatment sustains the treatment effect over many years time ("durability"). Second, we find that after the treatment is discontinued, it persists in generating an impact on the targeted behavior ("persistence") for as long as we observe the behavior – which is 15 to 23 months after the treatment is discontinued. Finally, we find that continued treatment generates a greater impact over time than a discontinued treatment. This suggests that the durability of the treatment effect is more than just persistence: that continued treatment exerts additive incremental influence on behavior. We hope that this work will be part of a wave of behavioral decision research which studies the intermediate- and long-term effects in field settings of behavioral interventions to improve societal well-being.

Cumulative Impact. The observations that this treatment produces persistence and durability have several implications for calculating the cumulative impact of this behavioral intervention – and other interventions that show persistence and durability, as well. Calculations of this type are of critical importance to policy-makers and managers since any calculation of cost effectiveness depends on having a sense of the cumulative impact of an intervention. For exactly that reason, the current research underscores the importance of policymakers and managers attending to intermediate- and long-run impacts of interventions before making decisions. First, when effects are persistent, the lifetime impact of a finite treatment period is substantially greater than the treatment effect measured during that finite period. Table 5 quantifies this for both experiments. Conservation during the joint treatment period is 525 kWh in Experiment 1 and 627 kWh in Experiment 2. During the following 15 and 23 months when we observe electricity use, those assigned to the discontinued condition conserve an additional 305 and 324 kWh in Experiments 1 and 2, respectively. This additional conservation increases the cumulative impact of treatment by 34 to 37%.

Given that we only monitored energy usage for a finite period of time after treatment was discontinued, and given that the persistence effect as seen in Figures 2 and 3 appears likely to survive beyond the two year period we observe, one might sensibly assume that the cumulative impact of the finite period of treatment is even greater than our data reflect. If one were to estimate this lifetime cumulative impact one would need to have a predicted rate of decay for the treatment effect. Allcott and Rogers (2012) estimate a linear decay rate (after controlling for seasonal differences in weather) for a similar treatment in a similar experiment conducted at a different location than the ones studied in the current two experiments. Using that specification, we estimate that the decay rate in Experiment 1 is 0.44 kWh/day per year (SE=0.09), more than twice the rate of 0.20 (SE=0.07) in Experiment 2. If these decay rates were to continue to hold into the future, it suggests that the total savings in each experiment would be on the order of twice as large as the effects during treatment. Of course, only time will tell whether or not the actual future decay rates are close to linear, and more generally what the cumulative savings will be.

Second, these results show that attributing the entire durability of the treatment effect to the continued treatment overstates the incremental impact of each successive Home Energy Report. This is because some of the energy use reduction observed after the joint treatment period among households assigned to the continued treatment condition is the result of the persistence of previous treatment, and not solely the result of the each additional treatment. The gap between the persistence effect and the durability effect is the incremental increase in treatment effect caused by continued administration of treatment after the joint treatment period. Table 5 shows that this incremental effect of continued treatment is only 31 to 49 percent of the energy use reduction among those in the continued condition after the joint treatment period. This calculation is of relevance to managers and policy makers who must decide whether or not to continue an existing intervention.

How is durability generated? Many factors might prevent durability from arising after a treatment is repeatedly administered. For example, as targets receive a treatment multiple times, they may become desensitized to it, they may attend to it less, and they may fail to react to it. This habituation may make treatments ineffective over time. But this is not what we observe: households decrease their energy usage as a result of repeatedly receiving the treatment over a period of years, and the result is not simply persistence. There are several features of the treatment that may contribute to this, not the least of which is that the descriptive social norms content is responsive to household behavior. In this way, the treatment – which does not change its aesthetic nor its psychological strategy – may be perceived as "unique" each time it is administered. The new data reflected in each report may reduce or prevent the habituation that one might expect of recipients after receiving the same treatment month after month. Future research can explore if this is one way that the treatment sustains attention, and thus maintains durability. We should note that durability is specifically not the result of the treatment automatizing behaviors like turning off lights, or increasing investments in energy efficient products. This is because those changes would be independent of continued administration of the treatment; they would be captured by our measure of persistence after treatment is discontinued.

How is persistence generated? An array of factors may contribute to the persistence of this treatment effect, which we classify into five categories. This taxonomy of how persistence can be generated is somewhat general to all behavioral interventions and so we illustrate each category with examples from other research in addition to how each category might contribute to the persistence studied in this manuscript. Though the categories are distinct, they almost certainly are interwoven, and the persistence of any given intervention could be the result of several of these pathways.

1. Set it and Forget it. One pathway through which behavioral interventions can show persistence is if the intervention induces participants to perform one-off behaviors that affect outcomes in the future, without further action. For example, interventions aimed at inducing people to enroll in 401(k) retirement savings plans by default enrolling new employees in the plans (Madrian and

Shea 2000) target a one-time behavior (enroll or not) that affects future outcomes performed by others on behalf of the target (deducting savings from one's paycheck over the course of many years). Once someone enrolls in such a plan a portion of all future paychecks is automatically redirected towards the retirement savings account, without any further action on the part of the target, and without psychologically changing the target. Similarly, purchasing an energy efficient air conditioner or weatherizing one's home involves a one-time decision that could lead to reduced energy consumption long after treatments are discontinued.

2. Memory. Another pathway through which behavioral interventions can generate persistence is if the intervention changes a target's memory content in specific ways that make the targeted behavior more likely. One route through which this might occur is by creating an association in memory between the performance environment and the targeted behavior. This is the psychological definition of a habit (Ouellette and Wood 1998), and these form through repeating a behavior in a specific environment. (The automaticity of psychological habits resembles Becker and Murphy's (1988) definition of habit as well). For example, when one of the authors enters the kitchen, he automatically opens the pantry door and collects a piece of chocolate – a persistent habit decades in the making. Or, when one leaves a room one may create a habit of turning off the lights such that whenever one leaves the room one automatically turns the lights off. Another route through which a behavioral intervention may affect persistence through memory through increasing the availability of some information such that it is more likely to be accessible to the target when the behavior is to be performed (Tversky and Kahneman 1974). For example, anti-smoking advertising that shows vivid images of people dying of lung cancer may increase the accessibility of lung cancer when the decision maker is deciding whether or not to smoke (Thrasher et al. 2012). Or, when purchasing light bulbs a consumer might remember his/her energy usage comparison and become more likely to purchase an energy efficient bulb.

3. Construal. Another pathway through which behavioral interventions can generate persistence is if the interventions change targets' construal of the information they encounter about themselves and the world. By changing how people perceive and interpret ambiguous information, interventions can change people's behaviors (Ross and Nisbett 1991). People are bombarded with information from the external world (performance feedback, social reactions, bills, etc.) and their internal worlds (their feelings, the attributions they make for success or failure, their heart rate, etc.). Behavioral interventions that modify this construal of themselves and the world effectively change the way people interpret and respond to (internal and external) events. Many of the most exciting behavioral interventions appear to leverage this pathway to persistent behavior change. For example, Walton and Cohen (2011) conducted a study involving a one-time intervention aiming to change how students construe social adversity on campus. This work built on previous research showing that feeling that one does not belong undermines motivation and academic performance (Walton and Cohen 2007). This intervention targeted African American students, a group that reports feeling socially isolated on many college campuses, with the aim increasing success in college. Outcome measures observed three years after the intervention showed improvements in grade-point average, as well as improvements in self-reported health, well-being, and number of doctor visits. Consistent with the construal interpretation, these researchers found that the persistent treatment effects appeared to be mediated by how students interpreted adversity in their social lives. (Other work in education mindsets could be classified in this category also, see Dweck 2007). For example, the treatment studied in this manuscript could have changed how households interpreted what a cold house in the summertime means. They could have come to interpret a cold house in the summertime as being an opulent extravagance rather than a pleasant luxury, thereby leading them to reduce their use of air conditioning.

4. Learning. Another pathway through which behavioral interventions can generate persistence is if the intervention allows targets' to learn about their preferences and to reduce ambiguity around behaviors. For example, inducing people to go to the gym for a few weeks may lead them to realize that the experience is not as unpleasant as they had expected, and therefore makes them more likely to exercise because of these revised expectations (Charness and Gneezy 2009). In the context of the current experiments, the treatment may have immediately induced households to try reducing their air conditioning usage just once. In the process of doing that they might have learned that a warmer house in the summertime is not as uncomfortable as they had expected. This learning allows them to modify their preferences so as to reflect what they have learned.

5. Rip currents. Another pathway through which behavioral interventions can generate persistence is through what we term "rip currents." A rip current is a channel of water in the ocean that runs perpendicular to the beach and carries anything that enters it very far into the ocean. If a person is just a foot out of the channel of water one is unaffected by the rip current; however if that person moves just one foot towards it that person could be carried miles out into the ocean by the rip current. In terms of behavioral interventions, one pathway through which persistence could be generated is by pushing people into the current of action in the world that will then engage them – and amplify the treatment – moving forward. This is very similar to Kurt Lewin's notion of "channel factors" (Lewin 1946). In get out the vote research, a common finding is that inducing people to vote in one election leads to greater turnout in later elections many years away (Gerber, Green, Shachar 2003; social pressure). One factor that may contribute to this is that once someone has voted in one election (and the public voter rolls show that this the person has voted), campaigns target that person differently and more intensively in future campaigns. In the context of the descriptive social norms treatment used in this manuscript, the treatment could have caused households to purchase an energy efficient product that resulted in their names being added to mailing lists for additional energy efficiency products or climate change advocacy, creating

increased opportunities for making investments in energy efficiency products, and increasing the number of energy conservation reminders a person encounters.

We are not able to assess the degree to which each of these pathways contributes to the persistence we observe in the two experiments reported in this manuscript. We can see that persistence mathematically depends on the rate at which a treatment effect decays once the treatment is discontinued, and we observe that the decay rate varies widely across experiments. As described above, the decay rate in Experiment 2 was less than half as rapid as the decay rate in Experiment 1. A third similar experiment involving the same treatment and design but implemented in a different site showed a decay rate that was barely one quarter the decay rate of Experiment 1 (0.12 kWh/day per year; see Allcott and Rogers 2012). Clearly decay rates vary substantially across settings for very similar treatments. Systematically studying what contributes to persistence is an important area for future research. Moreover, understanding how persistence occurs could generate strategies for enhancing the persistence (and, thus, the cumulative impact) of future interventions.

We study durability and persistence for only one treatment type (descriptive social norms messaging) targeting only one outcome (energy usage). Even though we replicate our main findings in two experiments, we certainly cannot generalize the findings to other types of interventions. In fact, as discussed above, despite the similarity in treatment across experiments, there is surprising variation in persistence across them. Future research will hopefully examine the long-term effects of other behavioral interventions in other domains, and, most critically, the factors that moderate these effects. We expect that the research like that reported in this manuscript will only grow in importance as behavioral science is increasingly called upon to inform solutions to vexing problems in the world.

REFERENCES

Allcott, H. 2011. Social norms and energy conservation. *J. Publ. Econom.* **95**(October 9-10) 1082-1095.

Allcott, H., M. Greenstone. 2012. Is there an energy gap? *J. Econom. Perspect.* **26**(1) 3-28.

Allcott, H., S. Mullainathan. 2010. Behavior and energy policy. *Science* **327**(5970).

Allcott, H., S. Mullainathan. 2012. External validity and partner selection bias. Working Paper No. 18373, The National Bureau of Economic Research.

Allcott, H., T. Rogers. 2012. The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *Unpublished manuscript.*

Ashraf, N., D. Karlan, W. Yin. 2006. Tying Odysseus to the mast: Evidence from a commitment savings product in the Phillipines. *Quart. J. Econom* **121**(2) 635-672.

Baumeister, R., K. Vohs, D, Funder. 2007. Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Sci.* **2**(4) 396-403.

Becker, G., K. Murphy. 1988. A Theory of Rational Addiction. *J. Political Econom.* **96**(4) 675-700.

Bertrand, M., S. Mullaninathan. 2004. Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *Amer. Econom. Review* **94**(4): 991-1013.

Bertrand, M., E. Duflo, S. Mullainathan. 2004. How much should we trust differences-in-differences estimates? *Quart. J. Econom.* **119**(1) 249-275.

Beshears, J., J. Choi, D. Laibson, B. Madrian. 2009. How Does Simplified Disclosure Affect Individuals' Mutual Fund Choices? Working Paper No. 14859. The National Bureau of Economic Research.

Cahill, K., R. Perera. 2008. Competitions and incentives for smoking cessation. *Cochrane Data. Systemetic Rev.* **3**.

Camerer, C., S. Issacharoff, G. Loewenstein, T. O'Donogue, M. Rabin. 2003. Regulation for conservatives: Behavioral economics and the case for 'Asymmetric Paternalism'. *University of Pennsylvania Law Review* **151** 1211.

Charness, G., U. Gneezy. 2009. Incentives to exercise. *Econometrica* **77**(3) 909-931.

Cialdini, R., L. Demaine, B. Sagarin, D. Barrett, K. Rhoads, P. Winter. 2006. Managing social norms for persuasive impact. *Social Influence* **1**(1) 3-15.

Cialdini, R., R. Reno, C. Kallgren. 1990. A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *J. Personality and Social Psychology* **58**(6) 1015-1026.

Dweck, C. 2007. *Mindset: The new psychology of success*. Random House Publishing, New York.

Ferraro, P.J., J.J. Miranda, M.K. Price. 2011. The Persistence of Treatment Effects with Norm-Based Policy Instruments: Evidence from a Randomized Environmental Policy Experiment. *Amer. Econom. Review: Papers and Proceedings* **101**(3) 318-322.

Frey, B., S. Meier. 2004. Social Comparisons and Pro-Social Behavior: Testing Conditional Cooperation in a Field Experiment. *Amer. Econom. Review*, **94**(5) 1717-1722.

Fryer, R.G., S.D.Levitt, J. List, S. Sadoff. 2012. Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment. Working Paper No. 18237. The National Bureau of Economic Research.

Gerber, A., D. Green, R. Shachar. 2003. Voting may be habit-forming: Evidence from a randomized field experiment. *Amer. J. Political Sci.* **47**(3) 540-550.

Gerber, A., T. Rogers. 2009. Descriptive social norms and motivation to vote: Everybody's voting and so should you. *J. Politics* **71** 1-14.

Gilovich, T., D. Griffin, D. Kahneman, eds. 2002. Heuristics and biases: The psychology of intuitive judgment. Cambridge University Press, Cambridge, U.K.

Gneezy, U., J. List. 2006. Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica* **74**(5) 1365-1384.

Goldstein, N., R. Cialdini, V. Griskevicius. 2008. A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *J. Consum. Research* **35** 472-482.

John, L., G. Loewenstein, A. Troxel, L. Norton, J. Fassbender, K. Volpp. 2011. Financial incentives for extended weight loss: A randomized, controlled trial. *J. Gen. Internal Medicine* **26**(6) 621-626.

Kahneman, D., A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* **47** 263-291.

Lewin, K. 1946. Behavior and development as a function of the total situation. L. Carmichael, ed. *Manual of Child Psychology*, John Wiley & Sons Inc., Hoboken, NJ, 791-844.

Madrian, B., D. Shea. 2000. The power of suggestion: Inertia in 401(k) participation and savings behavior. *Quar. J. Econom.* **116**(4) 1149-1187.

Nickerson, D., T. Rogers. 2010. Do you have a voting plan? Implementation intentions, voter turnout, and organic plan making. *Psychological Sci.* **21**(2) 194-199.

Ross, L., R. Nisbett. 1991. *The Person and the Situation: Perspectives on social psychology.* McGraw Hill, New York.

Nolan, J., W. Schultz, R. Cialdini, N. Goldstein, V. Griskevicius. 2008. Normative influence is underdetected. *Personality and Social Psychology Bulletin* **34** 913-923.

Ouellette, J., W. Wood. 1998. Habit and intention in everyday life: The multiple processes by which past behavior predicts future behavior. *Psychological Bulletin* **124**(1) 54-74.

Paluck, E. 2009. What's in a norm? Sources and processes of norm change. *J. Personality and Social Psychology* **96**(3) 594-600.

Reno, R., R. Cialdini, C. Kallgren. 1993. The transsituational influence of social norms. *J. Personality and Social Psychology* **64** 104-112.

Schultz, P., J. Nolan, R. Cialdini, N. Goldstein, V. Griskevicius. 2007. The constructive, destructive, and reconstructive power of social norms. *Psychological Sci.* **18**(5) 429-434.

Thaler, R., C. Sunstein. 2003. Libertarian Paternalism. *Amer. Econom. Review* **92**(2) 175-179.

Thaler, R., C. Sunstein. 2009. *Nudge: Improving decisions about health, wealth and happiness*. Penguin, New York.

Thrasher J., E. Arillo-Santillán, V. Villalobos, R. Pérez-Hernández, D. Hammond, J. Carter, E. Sebrié, R. Sansores, J. Regalado-Piñeda. 2012. Can pictorial warning labels on cigarette packages address smoking-related health disparities? Field experiments in Mexico to assess pictorial warning label content. *Cancer Causes Control* **23** 9-80.

Tversky, A., D. Kahneman. 1974. Judgement under uncertainty: Heuristics and biases. *Science* **185** 1124-1131.

Volpp, K., A. Troxel, M. Pauly, H. Glick, A. Puig, D. Asch, R. Galvin, J. Zhu, F. Wan, J. deGuzman, E. Corbett, J. Weiner, J. Audrain-McGovern. 2009. A randomized, controlled trial of financial incentives for smoking cessation. *New England J. of Medicine* **360**(February 12) 699-709.

Walton, J., G. Cohen. 2007. A question of belonging: Race, social fit, and achievement. *J. Personality and Social Psychology* **92**(1) 82-96.

Walton, J., G. Cohen. 2011. A brief social-belonging intervention improves academic health outcomes of minority students. *Science* **331**(6023) 1447-1451.

Yeager, D., G. Walton. 2011. Social-psychological interventions in education. *Rev. Educational Research* **81**(2) 267-301.

Table 1: Descriptive Statistics		
Experiment	**1**	**2**
Location	Upper Midwest	Southwest
Observations		
Total Number of Households	72,156	83,034
Continued Group	25,885	21,630
Discontinued Group	12,746	12,117
Control Group	33,525	49,287
Number of Observations	2,848,541	4,503,375
Balance		
Average Baseline Usage (kWh per day)	30.06	32.08
(Standard Deviation)	(16.65)	(15.58)
Treatment-Control Baseline Usage	0.024	-0.44
(Standard Error)	(0.12)	(0.51)
Continued - Discontinued Baseline Usage	-0.15	0.026
(Standard Error)	(0.18)	(0.19)
Attrition		
Percent of Treatment Group Opted Out	1.9%	2.6%
Percent of Accounts Inactive	15.2%	22.3%

	Table 2: Periods		
	Experiment:	**1**	**2**
Period	**Number**	**Begin Date**	
Baseline		October 1, 2007	April 1, 2006
Pre-Treatment	*0*	October 1, 2008	April 1, 2007
Early Joint Treatment Period	*1*	February 1, 2009	April 1, 2008
Late Joint Treatment Period	*2*	December 1, 2009	December 1, 2008
First 12 Months After Reports Discontinued	*3*	February 1, 2011	July 1, 2010
13-15 Months After Reports Discontinued	*4*	February 1, 2012	July 1, 2011
Remainder of Sample	*5*	None	October 1, 2011
Sample Ends		May 1, 2012	May 1, 2012

Table 3: Persistence and Durability		
Experiment	**1**	**2**
T · (Pre-Treatment)	-0.04	-0.01
	(0.06)	(0.06)
T · (Early Joint Treatment Period)	-0.49	-0.58
	(0.04)***	(0.09)***
T · (Late Joint Treatment Period)	-0.88	-0.84
	(0.05)***	(0.09)***
D · (First 12 Months After Reports Discontinued)	-0.73	-0.72
	(0.08)***	(0.12)***
D · (13-15 Months After Reports Discontinued)	-0.40	-0.67
	(0.11)***	(0.18)***
D · (Remainder of Sample)		-0.45
		(0.14)***
E · (First 12 Months After Reports Discontinued)	-0.98	-0.95
	(0.07)***	(0.11)***
E · (13-15 Months After Reports Discontinued)	-0.94	-1.11
	(0.09)***	(0.15)***
E · (Remainder of Sample)		-0.92
		(0.11)***
Month-by-Year Controls	Yes	Yes
Baseline Usage by Month-by-Year Controls	Yes	Yes
Number of Observations	2,659,622	4,411,214

*Notes: Independent variable is electricity consumption in kilowatt-hours per day. Robust standard errors, clustered by household. *, **, ***: Statistically significant with 90, 95, and 99 percent confidence, respectively.*

Table 4: Incremental Effects of Continued Treatment		
Experiment	**1**	**2**
T · (Pre-Treatment)	-0.04	-0.01
	(0.06)	(0.07)
T · (Early Joint Treatment Period)	-0.47	-0.57
	(0.04)***	(0.1)***
T · (Late Joint Treatment Period)	-0.86	-0.85
	(0.06)***	(0.09)***
T · (First 12 Months After Reports Discontinued)	-0.98	-0.95
	(0.07)***	(0.11)***
T · (13-15 Months After Reports Discontinued)	-0.94	-1.11
	(0.09)***	(0.15)***
T · (Remainder of Sample)		-0.92
		(0.11)***
D · (Pre-Treatment)	0.00	0.01
	(0.08)	(0.06)
D · (Early Joint Treatment Period)	-0.06	-0.03
	(0.06)	(0.07)
D · (Late Joint Treatment Period)	-0.07	0.04
	(0.08)	(0.08)
D · (First 12 Months After Reports Discontinued)	0.24	0.23
	(0.08)***	(0.09)**
D · (13-15 Months After Reports Discontinued)	0.54	0.44
	(0.11)***	(0.13)***
D · (Remainder of Sample)		0.47
		(0.11)***
Month-by-Year Controls	Yes	Yes
Baseline Usage by Month-by-Year Controls	Yes	Yes
Number of Observations	2,659,622	4,411,214

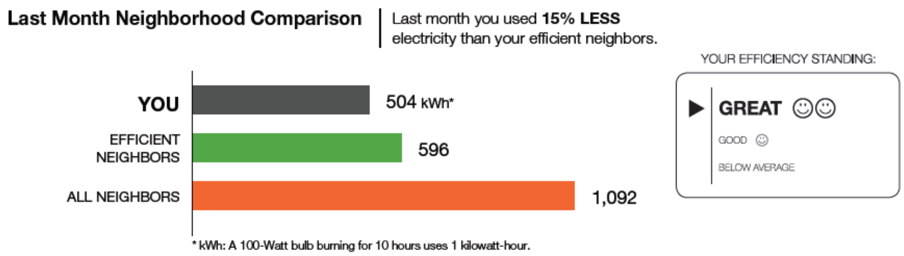
*Notes: Independent variable is electricity consumption in kilowatt-hours per day. Robust standard errors, clustered by household. *, **, ***: Statistically significant with 90, 95, and 99 percent confidence, respectively.*

Table 5: Total Electricity Conserved (Cumulative Impact)		
Experiment	**1**	**2**
Conservation During	525	627
Joint Treatment Period	(26)	(56)
Conservation from Discontinued Group	305	324
After Reports Discontinued	(32)	(47)
Conservation from Continued Group	444	253
After Reports Discontinued	(26)	(52)
Impact of Incremental Treatment	139	124
	(33)	(37)
Percent of Discontinued Group Savings	37%	34%
Incurred After Reports Discontinued		
Percent of Continued Group Savings	31%	49%
Attributable to Incremental Treatment		

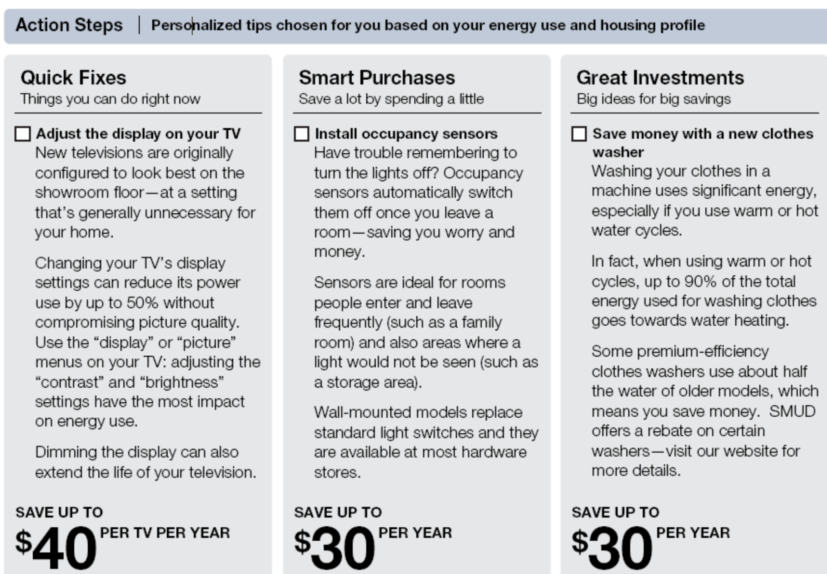
Notes: All figures in kilowatt-hours per household. Standard errors in parenthesis.

Figure 1: Opower Home Energy Report

(Front)



(Back)



Action Steps | **Personalized tips chosen for you based on your energy use and housing profile**

Quick Fixes
Things you can do right now

☐ **Adjust the display on your TV**
New televisions are originally configured to look best on the showroom floor—at a setting that's generally unnecessary for your home.

Changing your TV's display settings can reduce its power use by up to 50% without compromising picture quality. Use the "display" or "picture" menus on your TV: adjusting the "contrast" and "brightness" settings have the most impact on energy use.

Dimming the display can also extend the life of your television.

SAVE UP TO
$40 PER TV PER YEAR

Smart Purchases
Save a lot by spending a little

☐ **Install occupancy sensors**
Have trouble remembering to turn the lights off? Occupancy sensors automatically switch them off once you leave a room—saving you worry and money.

Sensors are ideal for rooms people enter and leave frequently (such as a family room) and also areas where a light would not be seen (such as a storage area).

Wall-mounted models replace standard light switches and they are available at most hardware stores.

SAVE UP TO
$30 PER YEAR

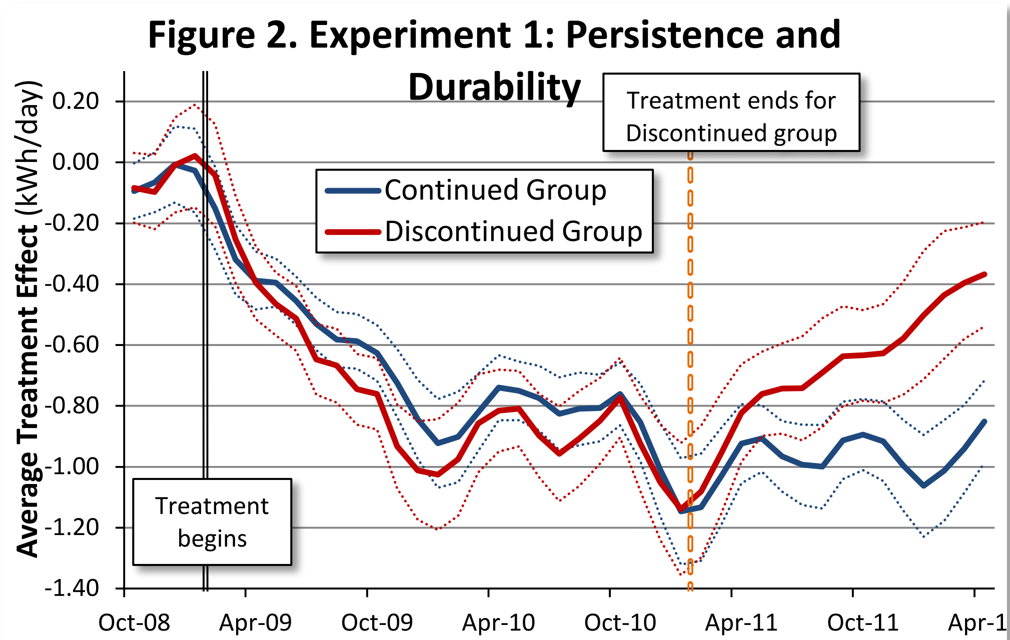
Great Investments
Big ideas for big savings

☐ **Save money with a new clothes washer**
Washing your clothes in a machine uses significant energy, especially if you use warm or hot water cycles.

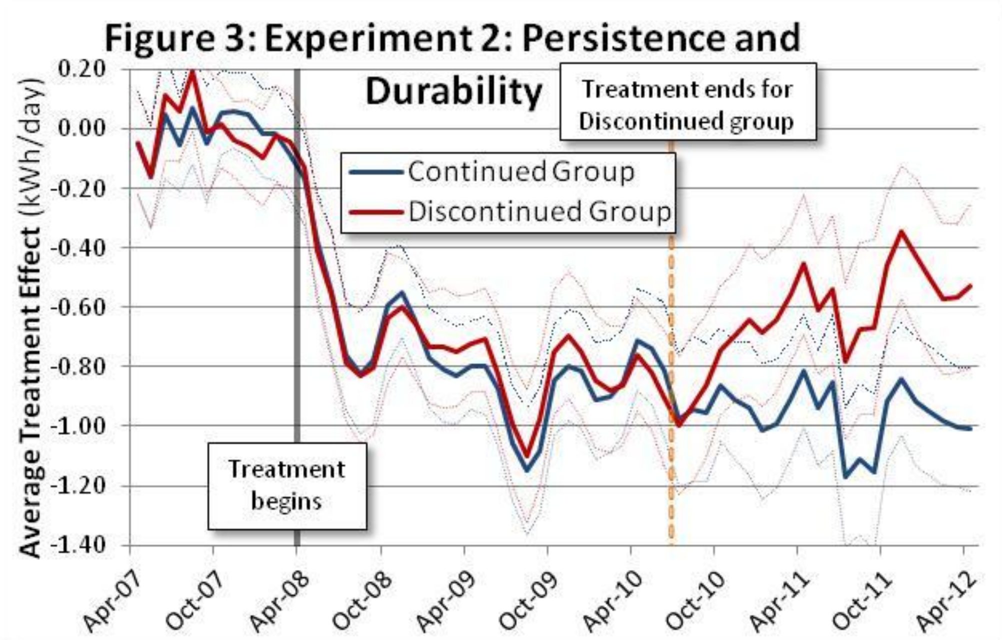
In fact, when using warm or hot cycles, up to 90% of the total energy used for washing clothes goes towards water heating.

Some premium-efficiency clothes washers use about half the water of older models, which means you save money. SMUD offers a rebate on certain washers—visit our website for more details.

SAVE UP TO
$30 PER YEAR



Note: This figure plots the ATEs for three month moving windows for those households assigned to the continued and discontinued conditions (compared to those in the control condition). The dotted lines represent 90 percent confidence intervals, with robust standard errors clustered by household.



Note: This figure plots the ATEs for three month moving windows for those households assigned to the continued and discontinued conditions (compared to those in the control condition). The dotted lines represent 90 percent confidence intervals, with robust standard errors clustered by household.